	Application No.	Applicant(s)
Notice of Allowability	10/736,273	HOULE, MICHAEL EDWARD
	Examiner	Art Unit
	MIRANDA LE	2167

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to *amendment filed on 01/04/08*.

2. ☒ The allowed claim(s) is/are *1-5, 8-9, now renumbered as 1-7*.

3. ☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a) ☒ All   b) ☐ Some*   c) ☐ None of the:

    1. ☒ Certified copies of the priority documents have been received.

    2. ☐ Certified copies of the priority documents have been received in Application No. _____ .

    3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

    * Certified copies not received: _____ .

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application. **THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.**

4. ☐ A SUBSTITUTE OATH OR DECLARATION must be submitted. Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL PATENT APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.

    (a) ☐ including changes required by the Notice of Draftsperson's Patent Drawing Review (PTO-948) attached

       1) ☐ hereto or 2) ☐ to Paper No./Mail Date _____ .

    (b) ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____ .

Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)

2. ☐ Notice of Draftperson's Patent Drawing Review (PTO-948)

3. ☐ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____

4. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material

5. ☐ Notice of Informal Patent Application

6. ☒ Interview Summary (PTO-413), Paper No./Mail Date _____ .

7. ☒ Examiner's Amendment/Comment

8. ☐ Examiner's Statement of Reasons for Allowance

9. ☐ Other _____ .

DETAILED ACTION

1.   This communication is responsive to Amendment filed 01/04/2008.

Claims 1-17 are pending in this application. Claims 1, 4, 8 are independent

claims. In the Amendment, claims 6, 7 have been cancelled, claims 1, 2, 4, 8 have been

amended.

Election/Restrictions

2.   This application is in condition for allowance except for the presence of claims

10-17 directed to be drawn as non-elected without traverse. Accordingly, claims 10-17

have been cancelled.

EXAMINER'S AMENDMENT

3.   An examiner's amendment to the record appears below. Should the changes

and/or additions be unacceptable to applicant, an amendment may be filed as provided by

37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no

later than the payment of the issue fee.

Authorization for this examiner's amendment was given in a telephone interview

with Applicant's representative, Mr. David Bowls, on February 29, 2008.

The application has been amended as follows:

Claim 1 has been amended as:

A computer system for generating data structures for information retrieval of documents

stored in a database, said documents being stored as document-keyword vectors

generated from a predetermined keyword list, and said document-keyword vectors

forming nodes of a hierarchical structure imposed upon said documents, said computer

system comprising:

a processor having accessed to the database;

a document-keyword matrix generation subsystem;

a neighborhood patch generation subsystem for generating groups of nodes

having similarities as determined using a search structure, said neighborhood patch

generation subsystem including a subsystem for generating a spatial approximation

sample hierarchy structure upon said document-keyword vectors and a patch defining

subsystem for creating patch relationships among said nodes with respect to a metric

distance between nodes;

a query vector generation subsystem accepting search conditions and query

keywords, generating a corresponding query vector, and storing the generated query

vector;

an intra-patch confidence and inter-patch confidence determination subsystem for

every element of the database, the spatial approximation sample hierarchy structure

computing a neighborhood patch consisting of a list of those database elements most

similar to it for computing inter-patch confidence values between patches and intra-patch

confidence values;.

a self confidence determining subsystem for (a) computing a list of self

confidence values, for every stored patch, (b) computing relative self confidence values,

and (c) thereafter using the relative self confidence values to determine a size of a best

subset of each patch to serve as a cluster candidate;

a cluster estimation subsystem for generating cluster data of said document-keyword-vectors using said similarities of patches wherein the cluster estimation subsystem selects said patches depending on-intra-patch confidence values to represent clusters of said document keyword vectors, estimate the sizes of said patches, and generate cluster data of document keyword vectors using similarities of the patches;

a redundant cluster elimination subsystem for using inner patch confidence values to eliminate redundant cluster candidates; and

a display subsystem for displaying on screen said estimated clusters together with confidence relations between said clusters and hierarchical information pertaining to cluster size.

- **Claim 4 has been amended as:**

A method for generating data structures for information retrieval of documents stored in a database, said documents being stored as document-keyword vectors generated from a predetermined keyword list, and said document-keyword vectors forming nodes of a hierarchical structure imposed upon said documents, said method comprising the step of:

generating a hierarchical structure upon said document-keyword vectors and storing hierarchy data in an adequate storage area;

generating neighborhood patches of nodes having similarities as determined using levels of the hierarchical structure, and storing said patches in an adequate storage area;

generating groups of nodes having similarities as determined using a search

structure, including generating a spatial approximation sample hierarchy structure upon

said document-keyword vectors and creating patch relationships among said nodes with

respect to a metric distance between nodes;

determining inter-patch confidence values between patches and intra-patch

confidence values;

determining an intra-patch confidence and inter-patch confidence for every

element of the database, comprising utilizing the spatial approximation sample hierarchy

structure to compute a neighborhood patch consisting of a list of those database elements

most similar to it and computing inter-patch confidence values between patches and intra-

patch confidence values;

determining self confidence values to determine a size of a best subset of each

patch to serve as a cluster candidate by the steps of (a) computing a list of self confidence

values, for every stored patch, (b) computing relative self confidence values, and (c)

thereafter using the relative self confidence values to determine the size of a best subset

of each patch to serve as a cluster candidate;

invoking said hierarchy data and said patches to compute inter-patch confidence

values between said patches and intra-patch confidence values, and storing said values as

corresponding lists in an adequate storage area;

estimating the sizes of said patches, and generating cluster data of document-

keyword vectors using similarities of the patches, selecting said patches depending on

said inter-patch confidence values and said intra-patch confidence values to represent

clusters of said document-keyword vectors: and

using inner patch confidence values to eliminate redundant cluster candidates; and

displaying on screen said estimated clusters together with confidence relations

between said clusters and hierarchical information pertaining to cluster size.

- **Claim 8 has been amended as:**

A computer-readable storage medium storing a program for making a computer system

execute a method for generating data structures for information retrieval of documents

stored in a database, said documents being stored as document-keyword vectors

generated from a predetermined keyword list, and said document-keyword vectors

forming nodes of a hierarchical structure imposed upon said documents, said program

making said computer system execute the steps of:

accepting search conditions and query keywords, generating a corresponding

query vector, and storing the generated query vector;

generating a hierarchical structure upon said document-keyword vectors and storing

hierarchy data in an adequate storage area;

generating neighborhood patches consisting of nodes having similarities as

determined using levels of the hierarchical structure, and storing said patch list in an

adequate storage area;

generating groups of nodes having similarities as determined using a search

structure, including generating a spatial approximation sample hierarchy structure upon

said document-keyword vectors and creating patch relationships among said nodes with

respect to a metric distance between nodes;

determining an intra-patch confidence and inter-patch confidence for every

element of the database, comprising utilizing the spatial approximation sample hierarchy

structure to compute a neighborhood patch consisting of a list of those database elements

most similar to it and computing inter-patch confidence values between patches and inter-

patch confidence values;

determining self confidence values to determine a size of a best subset of each

patch to serve as a cluster candidate by the steps of (a) computing a list of self confidence

values, for every stored patch, (b) computing relative self confidence values, and (c)

thereafter using the relative self confidence values to determine the size of a best subset

of each patch to serve as a cluster candidate;

invoking said hierarchy data and said patches to compute inter-patch confidence

values between said patches and intra-patch confidence values, and storing said values as

corresponding lists in an adequate storage area;

selecting said patches depending on said inter-patch confidence values and said

intra-patch confidence values to represent clusters of said document-keyword vectors;

using inner patch confidence values to eliminate redundant cluster candidates; and

displaying on screen said estimated clusters together with confidence relations

between said clusters and hierarchical information pertaining to cluster size.

- **Claim 9 has been amended as:**

The computer readable storage medium according to claim 8, further comprising

the step of estimating sizes of said clusters depending on said intra-patch confidence

values.

Reasons for Allowance

3.   Claims 1-5, 8-9 are allowed, now renumbered as 1-7.

4.   The following is a statement of reasons for the indication of allowable subject

matter:

The present invention is directed to a system and method for information retrieval

and data mining of text databases, using shared neighbor information to determine query

clusters. The clustering method assesses the level of mutual association between a query

element (which may or may not be an element of the data set) and its neighborhood

within the data set. The association between two·elements is considered strong when the

elements have a large proportion of their nearest neighbors in common. Methods are

based on the new and original concepts of inter-cluster association confidence and intra-

cluster association self-confidence.

Claim 1 recites, or similarly recites, in combination with the remaining elements,

a computer system comprising:

**the spatial approximation sample hierarchy structure computing a neighborhood
patch consisting of a list of those database elements most similar to it;**

**a self confidence determining subsystem for computing a list of self confidence
values for every stored patch, computing relative self confidence values, and thereafter
using the relative self confidence values to determine a size of a best subset of each patch
to serve as a cluster candidate;**

a redundant cluster elimination subsystem for using inner patch confidence values to eliminate redundant cluster candidates.

The closest prior art, et al. Tang et al. (U.S. Patent No. 6,636,849), shows a substantially similar data search employing metric spaces method wherein application of Tang may include textual or byte-based searches, literature search based on lists of keywords, and vector and matrix based indexing and searching (Abstract). While Tang discloses multigrid (i.e. group of nodes) search tree to find exact or approximate or homologous matches for a search query (i.e. creating patch relationships among said nodes with respect to a metric distance between nodes); and Gilmour et al. (U.S. Patent No. 6,377,949) teaches assigning a confidence level to a term within an electronic document or a second quantitative indicator, Tang et al. and Gilmour et al., singularly or in combination, still fail to anticipate or render the above cited limitations obvious.

Claim 4 recites, or similarly recites, in combination with the remaining elements, the method comprising the steps of:

the spatial approximation sample hierarchy structure computing a neighborhood patch consisting of a list of those database elements most similar to it;

a self confidence determining subsystem for computing a list of self confidence values for every stored patch, computing relative self confidence values, and thereafter using the relative self confidence values to determine a size of a best subset of each patch to serve as a cluster candidate;

a redundant cluster elimination subsystem for using inner patch confidence values to eliminate redundant cluster candidates.

The closest prior art, et al. Tang et al. (U.S. Patent No. 6,636,849), shows a substantially similar data search employing metric spaces method wherein application of Tang may include textual or byte-based searches, literature search based on lists of keywords, and vector and matrix based indexing and searching (Abstract). While Tang

discloses multigrid (i.e. group of nodes) search tree to find exact or approximate or

homologous matches for a search query (i.e. creating patch relationships among said

nodes with respect to a metric distance between nodes); and Gilmour et al. (U.S. Patent

No. 6,377,949) teaches assigning a confidence level to a term within an electronic

document or a second quantitative indicator, Tang et al. and Gilmour et al., singularly or

in combination, still fail to anticipate or render the above cited limitations obvious.

Claim 8 recites, or similarly recites, in combination with the remaining elements,

the steps of:

**the spatial approximation sample hierarchy structure computing a neighborhood
patch consisting of a list of those database elements most similar to it;**

**a self confidence determining subsystem for computing a list of self confidence
values for every stored patch, computing relative self confidence values, and thereafter
using the relative self confidence values to determine a size of a best subset of each patch
to serve as a cluster candidate;**

**a redundant cluster elimination subsystem for using inner patch confidence values
to eliminate redundant cluster candidates.**

The closest prior art, et al. Tang et al. (U.S. Patent No. 6,636,849), shows a

substantially similar data search employing metric spaces method wherein application of

Tang may include textual or byte-based searches, literature search based on lists of

keywords, and vector and matrix based indexing and searching (Abstract). While Tang

discloses multigrid (i.e. group of nodes) search tree to find exact or approximate or

homologous matches for a search query (i.e. creating patch relationships among said

nodes with respect to a metric distance between nodes), and Gilmour et al. (U.S. Patent

No. 6,377,949) teaches assigning a confidence level to a term within an electronic

document or a second quantitative indicator, Tang et al. and Gilmour et al., singularly or

in combination, still fail to anticipate or render the above cited limitations obvious.

5.   Any comments considered necessary by applicant must be submitted no later than

the payment of the issue fee and, to avoid processing delays, should preferably

accompany the issue fee. Such submissions should be clearly labeled "Comments on

Statement of Reasons for Allowance".

Conclusion

6.   Any inquiry concerning this communication or earlier communications from the

examiner should be directed to Miranda Le whose telephone number is (571) 272-4112.

The examiner can normally be reached on Monday through Friday from 8:30 AM to 5:00

PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, John R. Cottingham, can be reached on (571) 272-7079. The fax number to

this Art Unit is 571-273-8300.

Any inquiry of a general nature or relating to the status of this application should

be directed to the Group receptionist whose telephone number is (571) 272-2100.

Information regarding the status of an application may be obtained from the

Patent Application Information Retrieval (PAIR) system. Status information for

published applications may be obtained from either Private PAIR or Public PAIR. Status

information for unpublished applications is available through Private PAIR only. For

more information about the PAIR system, see http://pair-direct.uspto.gov. Should you

have questions on access to the Private PAIR system, contact the Electronic Business

Center (EBC) at 866-217-9197 (toll-free).

Miranda Le
February 29, 2008